The evolutionary genomics of anthroponosis in Cryptosporidium

Nader, Johanna ; Mathers, Thomas; Ward, Ben; Pachebat, Justin Alexander; Swain, Martin; Robinson, Guy; Chalmers, Rachel M.; Hunter, Paul R.; van Oosterhout, Cock; Tyler, Kevin

Published in:
Nature Microbiology

DOI:
10.1038/s41564-019-0377-x

Publication date:
2019

Citation for published version (APA):
Nader, J., Mathers, T., Ward, B., Pachebat, J. A., Swain, M., Robinson, G., Chalmers, R. M., Hunter, P. R., van Oosterhout, C., & Tyler, K. (2019). The evolutionary genomics of anthroponosis in Cryptosporidium. *Nature Microbiology*, *4*(N/A), 826-836. https://doi.org/10.1038/s41564-019-0377-x

Title: Evolutionary genomics of anthroponosis in *Cryptosporidium*

Authors and Affiliations

Johanna L. Nader[1,2], Thomas C. Mathers[3], Ben J. Ward[3,4], Justin A. Pachebat[5], Martin T. Swain[5], Guy Robinson[6,7], Rachel M. Chalmers[6,7], Paul R. Hunter[1], Cock van Oosterhout[4*], Kevin M. Tyler[1*]

[1]Biomedical Research Centre, Norwich Medical School, University of East Anglia, Norwich, United Kingdom
[2]Department of Genetics and Bioinformatics, Division of Health Data and Digitalisation, Norwegian Institute of Public Health, Oslo, Norway
[3]Earlham Institute, Norwich Research Park, Norwich, United Kingdom
[4]School of Environmental Sciences, Norwich Research Park, University of East Anglia, United Kingdom
[5]Institute of Biological, Environmental & Rural Sciences, Aberystwyth University, Aberystwyth, United Kingdom
[6]*Cryptosporidium* Reference Unit, Public Health Wales Microbiology, Singleton Hospital, Swansea, United Kingdom
[7]Swansea University Medical School, Singleton Park, Swansea, United Kingdom

*Contributed equally to the work

Corresponding authors:
Email: johanna.nader@fhi.no
Telephone: +47 41221727
Address: Norwegian Institute of Public Health, Postbox 4404, Nydalen 0403, Oslo, Norway

Email: c.van-oosterhout@uea.ac.uk
Telephone: +44 1603 592921
Address: School of Environmental Sciences, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK

Abstract

Human cryptosporidiosis is the leading protozoan cause of diarrhoeal mortality worldwide, and a preponderance of infections is caused by *Cryptosporidium hominis* and *C. parvum*. Both species consist of several subtypes with distinct geographic distributions and host preferences (i.e. generalist zoonotic and specialist anthroponotic subtypes). The evolutionary processes driving the adaptation to human host, and the population structure remain unknown. In this study, we analyse 21 whole genome sequences to elucidate the evolution of anthroponosis. We show that *C. parvum* splits into two subclades, and that the specialist anthroponotic subtype IIc-a shares a subset of loci with *C. hominis* that are undergoing rapid convergent evolution driven by positive selection. Subtype IIc-a also has an elevated level of insertion-deletion (indel) mutations in the peri-telomeric genes, which is characteristic also for other specialist subtypes. Genetic exchange between subtypes plays a prominent role throughout the evolution of *Cryptosporidium*. Interestingly, recombinant regions are enriched for positively selected genes and potential virulence factors, which indicates adaptive introgression. Analysis of 467 gp60 sequences collected across the world shows that the population genetic structure differs markedly between the main zoonotic subtype (isolation-by-distance) and the anthroponotic subtype (admixed population structure). Finally, we show that introgression between the four anthroponotic *Cryptosporidium* subtypes and species included in this study has occurred recently, probably within the past millennium.

Introduction

Diarrhoeal pathogens cause more mortality than malaria, measles, and AIDS combined[1] and globally, for children under five, *Cryptosporidium* is the leading, vaccine non-preventable cause of diarrhoeal morbidity and mortality[2]. The zoonotic *Cryptosporidium parvum* and the anthroponotic *Cryptosporidium hominis* account for a vast majority of such cases. *C. hominis* and *C. parvum* have consistently been reported as exhibiting a high average global consensus of ~95-97% nucleotide identities[3,4]; yet, the genetic basis for the difference in host range has remained unexplained, and our understanding of host adaptation is confounded by the existence of anthroponotic *C. parvum* isolates (Supplementary Fig. S1). The relatively high level of genomic conservation between these species could be explained by similarity in selection pressures experienced by these parasites that is irrespective of their hosts. For example, *Plasmodium berghei* requires two-thirds of genes for optimal growth during a single stage of its complex life cycle[5]. Alternatively, hybridization amongst isolates of *Cryptosporidium* species could lead to genetic introgression that homogenizes sequence variation. For example, some "generalist" plant pathogens such as the oomycete *Albugo candida* have a huge host range consisting of hundreds of plant species that are parasitized by host-specific subtypes[6]. This pathogen suppresses the immune response of the host plant, enabling hybridization between different subtypes leading to genetic introgression that is thought to fuel the coevolutionary arms race[3,8] . Similarly, in the mosaic-like *Toxoplasma gondii* genomes there are conserved chromosomal haploblocks which are shared across otherwise diverged clades[7].

The ~9.14Mbp *Cryptosporidium* genome comprises 8 chromosomes ranging in size from 0.88 to 1.34Mbp, and has a highly compact coding sequence composition (73.2-77.6%)[8]. Genomic comparisons between the original *C. parvum* Iowa[9] and *C. hominis* TU502[10] reference genomes currently provide an overview of chromosome-wide hotspots for single nucleotide polymorphisms (SNPs), selective pressures, and species-specific genes and duplication events[4,11]. These studies revealed peri-telomeric clustering of hyper-polymorphism and identified several putative virulence factors. Attempts to correlate

genomic changes with phenotypic expression identified only a few shared SNPs between the anthroponotic *C. parvum* and *C. hominis*[12]. Whole genome comparisons found genome-wide incongruence and significant sequence insertion and deletion (indels) events between *C. hominis* and *C. parvum*[13], and recombination at the hypervariable gp60 subtyping locus[14]. Expanding cross-comparisons to include multiple whole genome sequences (WGS) across a range of anthroponotic and zoonotic *C. parvum* and *C. hominis* strains will help to explore these phenotype-associated features, and understand the evolution of human-infective strains.

Here, we have conducted a phylogenetic comparison of 21 WGS, including 11 previously unpublished *Cryptosporidium* genome sequences (Table S1). In addition, we characterise the global distribution of *Cryptosporidium* species and subtypes, summarising the data of 743 peer-reviewed publications of cases in a total of 126 countries that used the gp60 locus for species identification and subtyping. We describe the evolutionary genomic changes of this pathogen during its association with its human host and host-range specialisation, and we estimate divergence times for the primary anthroponotic lineages. Our analyses provide a revised evolutionary scenario supporting the more recent emergence of a previously cryptic, phylogenetically-distinct anthroponotic *Cryptosporidium parvum anthroponosum* sub-species.

Results

A phylogenetic analysis of 61 neutrally-evolving coding loci across 21 *Cryptosporidium* isolates reveals the evolutionary history of human-infective taxa and identifies two discrete *C. parvum* lineages with distinct host associations, namely *C. p. parvum* (zoonotic) and *C. p. anthroponosum* (anthroponotic) (Fig. 1a; Fig. S1)[13]. Primary human-infective isolates[15] *C. hominis* and *C. parvum* form a distinct superclade with zoonotic *C. cuniculus*, a recently-identified cause of human outbreaks[16,17]. This superclade is genetically distinct from other zoonotic human-infectious *Cryptosporidium* species (*C. meleagridis*[18], *C. viatorum*[19], *C. ubiquitum*[20], *C. baileyi*[21] and *C. muris*[22]; Fig. 1a; Fig. S2; absolute divergence (d_{xy}) = 0.083 – 0.478). Within the superclade, limited genetic divergence between *C. hominis* and *C. parvum* (d_{xy} = 0.031) illustrates the recent origins of these taxa. Finally, the concatenated phylogeny provides a preliminary genotypic association between phenotypically-diverse *C. parvum* strains. Based on the host ranges of a total of 1331 isolates, *C. p. anthroponosum* UKP15 (subtype IIc-a) is almost exclusively found in humans (92.2%), whereas *C. p. parvum* UKP6 and UKP8 (subtypes IIa and IId, respectively) are more often found in ruminants than in humans (Fig. 1S). These zoonotic subtypes (UKP6 and UKP8) split off into a unique sister group (*C. p. parvum*) within the *C. parvum* clade, distinct from the anthroponotic subtype (*C. p. anthroponosum*). This switch in host association is associated with surprisingly low levels of genetic divergence (d_{xy} = 0.002), suggesting it happened recently.

Next, we undertook a meta-analysis to establish the distribution and population genetics of these *Cryptosporidium* species and subtypes based on gp60 genotyping, summarising the data of 743 peer-reviewed publications of cases in a total of 126 countries worldwide published between 2000 and 2017. The anthroponotic species *C. hominis* and *C. p. anthroponosum* are relatively more prevalent in resource poor countries (Fig. 1b,c). In contrast, the zoonotic *C. p. parvum* dominates in North America, Europe, parts of the Middle East and Australia. Even though *C. p. anthroponosum* is less prevalent in Europe (17%; 22 out of 128 cases), the mean nucleotide diversity at gp60 is significantly higher than that of *C. p. parvum* (π = 0.02954 vs. 0.00327, respectively) (Mann-Whitney test: W= 430412; $p < 10^{-5}$) (Fig. 1d). The population genetic structure differs significantly between *C. p. anthroponosum* and *C. p. parvum* (GLM:

$F_{1,79} = 47.34$, $p < 0.0001$), with *C. p. parvum* showing a strong isolation-by-distance signal, whereas there is no geographic population genetic structure for *C. p. anthroponosum* (Fig. 1e; Tables S2, S3). In Europe, *C. p. parvum* forms a geographically-structured population which shows significant isolation-by-distance (Fig. 1f,g). This suggests that gene flow within Europe shapes the genetic differentiation (F_{st}) of *C. p. parvum*, and that this pathogen is transmitted between European countries. In contrast, the high nucleotide diversity and lack of geographic structuring implies that *C. p. anthroponosum* may be introduced in Europe from genetically diverged source populations. The population genetic structure of both species is also different when analysed across a global-scale, with network analysis revealing significant sub-structuring of global populations of *C. p. parvum*, but not of *C. p. anthroponosum* (Fig. 1g,h).

Nucleotide divergence between *C. p. parvum* and *C. p. anthroponosum* is driven partly by positive selection, as evidenced by the relatively elevated ratio of Ka/Ks (> 1.0) for 44 loci (Fig. 2a; Table S4). The Ka/Ks ratio between the *C. parvum* subspecies is comparable to the Ka/Ks ratio of *C. p. parvum* and *C. hominis* comparison, and significantly higher than the Ka/Ks ratio of comparisons between other *C. p. parvum* subtypes (Fig. 2b). The signature of adaptive evolution is most apparent in the peri-telomeric genes (Fig. S4). Furthermore, frameshift-causing indels also underpin protein divergence in 130 (55.6%) and 24 (53.3%) variable *C. hominis* and *C. p. anthroponosum* amino acid coding sequences, respectively (Table S5, S6). When accounting for the size of the different chromosomal regions, indels are significantly more common in the peri-telomeric and subtelomeric regions than elsewhere in the genome (Chi-sq. test: $X^2 = 257.71$, df = 2, $p = 1.09 \times 10^{-56}$) (Fig. 2c). Genes encoding for extracellular proteins show a significantly stronger signal of positive selection than genes with a cytoplasmic protein localization (Mann-Whitney test: W = 842985, p = 0.0182) (Fig. 2d; S5), consistent with adaptations/specialisation to the human host.

Besides nucleotide substitutions and indels, genetic introgression also appears to play a prominent role in the adaptive evolution of *Cryptosporidium*. To investigate genome-wide patterns of divergence between *Cryptosporidium* lineages we aligned reads from 16 isolates to the *C. parvum* Iowa reference genome[9]. Principle component analysis based on a set high quality SNPs supports the sub-species assignments of zoonotic *C. p. parvum* and anthroponotic *C. p. anthroponosum* (Fig. 3a). Surprisingly, one sample (UKP16), identified as *C. p. parvum* based on phylogenetic analysis of 61 single copy conserved genes (Fig. 1a), appears to be highly differentiated based on genome wide SNPs (Fig. 3a). To further investigate the evolutionary history of this sample we generated phylogenetic trees in 50 SNP windows across the genome. The consensus topology of these genomic windows is shown as a "cloudogram" (Fig. 3b), which matches the concatenated analysis of conserved protein coding genes (Fig. 1a), with UKP16 most closely related to *C. p. parvum* isolates. However, many alternative topologies are also observed, indicating potential recombination between lineages (Fig. 3b). We used topology weighting[23] to visualise the distribution of topologies across the genome, focusing on evolutionary relationships between UKP16, *C. p. parvum* isolates and *C. p. anthroponosum* isolates (Fig. 3c). This analysis revealed a large region in chromosome 8 (~500 - 650Kb) where UKP16 has a sister relationship to *C. p. parvum* isolates and *C. p. anthroponosum* isolates (topo1; Fig. 3c and d). Intriguingly, this appears to be due introgression into the UKP16 genome from a highly divergent, and as yet unsampled, lineage. We draw this conclusion because the absolute divergence (d_{xy}) between UKP16 and both *C. p. anthroponosum* and *C. p. parvum* is elevated in this region, whereas divergence between *C. p. anthroponosum* and *C. p. parvum* is similar to the rest of the chromosome (Fig. 3e).

Next, we conducted a detailed analysis of genetic introgression, studying two *C. parvum parvum* isolates (UKP6 and UKP16), one *C. parvum anthroponosum* isolate (UKP15), and one *C. hominis* isolate (UKH1). A total of 104 unique recombination events are detected across these four whole genome sequences (Fig 4a; Table S7). Many recombination events involve an unknown parental sequence (i.e. donor), which is consistent with our findings for the UKP16 sample, where we identified an introgressed genomic segment from a diverged lineage (see above). These results highlight that genetic exchange is widespread across *Cryptosporidium* species. The distribution of recombination events varies markedly across chromosomes, with a disproportionately higher number of individual events detected in chromosome 6 (25.9% of total events), and a disproportionately lower number of events in chromosomes 3, 5, and 7 (Fig. S6). Another consequence of introgression is that the coalescence time between different subtypes can vary markedly within and across chromosomes, ranging from an estimated 776 to 146,415 generations ago (Table S7). Furthermore, many recombination events are detected in the peri-telomeric genes (Fig. 4a). Interestingly, of the 44 genes that appear to be under positive selection (Ka/Ks>1; see Fig. 2a), no less than 17 (38.64%) are affected by recombination. This is significantly higher than the 6.57% of genes (237 out of 3607 genes) affected by recombination that are neutrally evolving or under purifying selection (Ka/Ks<1) (Chi-square test: $X^2 = 54.51$, df = 1, p = 1.55×10^{-13}). In addition, a significantly greater number of recombination events is observed in *C. p. anthroponosum* (n=39) than in *C. hominis* (n=7) (binomial test: p = 3.12×10^{-7}) and *C. p. parvum* (n=17) (binomial test: p = 0.011) (Table S7). These analyses suggest that the genetic exchange between diverged lineages is unlikely to be a neutral process and may be fuelling adaptation in anthroponotic lineages of *Cryptosporidium*.

Finally, we estimate the divergence dates to provide the first chronological description for genetic introgression between human-infective *Cryptosporidium* spp. (Fig. 4b). The majority of introgression events between *C. p. parvum* and *C. p. anthroponosum* strains are estimated to have taken place at approximately 10-15 thousand generations ago (TGA). Only circa 6.8% of all genetic exchanges are introgression events into the *C. hominis* genome, and as expected, these events are more ancient (i.e. ~75-150 TGA). To translate generation time into years and estimate the age of the introgression events, we assume a generation time of between 48 and 96 hours[24,25], and a steady rate of transmission within host populations. The following estimates should be considered minimum estimates of divergence times because *Cryptosporidium* may be dormant outside the host. We estimate that the zoonotic *C. p. parvum* and the anthroponotic *C. p. anthroponosum* strains are likely to have recombined between 55-164 years ago, whereas we estimate that introgression events between *C. hominis* and *C. parvum* occurred between 410-1096 years ago (Fig. 4b). We show that despite genetic adaptation to specific hosts, diverged *Cryptosporidium* (sub)species continue to exchange genetic information through hybridisation within the last millennium, and that such exchange does not appear to be selectively neutral.

Discussion

Cryptosporidium is an apicomplexan parasite that can cause debilitating gastrointestinal illness in animals and humans worldwide. In order to better understand the biology of this parasite, we conducted an analysis to describe the population structuring based on 467 sequences of a highly-polymorphic locus (gp60), and we study the evolution of this parasite using 16 whole genome sequences. We demonstrate here that *C. parvum* consists of two subspecies with distinct host associations, namely *C. p. parvum* (zoonotic) and *C. p. anthroponosum* (anthroponotic) that have diverged recently. Nevertheless, the population

genetic structure differs significantly between both subspecies, with *C. p. parvum* showing a strong isolation-by-distance signal, whilst there is no clear geographic structure for *C. p. anthroponosum*. Besides the apparent differences in drift and gene flow, the divergence of both subspecies is also driven by positive selection, and the signature of adaptive evolution is comparable to that of *C. p. parvum* and *C. hominis*. Perhaps most remarkably, hybridisation has frequently led to the genetic introgression between these (sub)species. Given that such exchanges appear to be associated in particular to genes under positive selection, we believe that hybridisation plays an important role throughout the evolution of these parasites. Next, we describe *Cryptosporidium* biology with the aim to interpret and explain the population genetic and evolutionary genetic findings, placing them into the context of recent whole genome studies of other pathogens.

Our population genetic analysis detected remarkable differences between *C. p. anthroponosum* and *C. p. parvum*, both in their population genetic structure, as well as their levels of nucleotide diversity. *C. p. parvum* can cause neonatal enteritis (scour) predominantly in pre-weaned calves[26]. Given that such calves are able to produce circa 100,000 oocysts per gram of faeces, they are thought to be the primary source of subsequent infections[27]. Movement of such young animals has therefore been highly restricted by the European Union[28,29]. Adult cattle tend to be asymptomatic and shed fewer oocysts, and consequently, they are believed to be minor transmission vectors. Furthermore, long distance translocation of cattle is rare compared to human migration; just 42,515 cattle were exported to the EU from the UK[30] whereas 70.8 million overseas visits were made by UK residents in 2016[31]. Consequently, in cattle *C. p. parvum* mediated scour is unlikely to be spread by long distance migration via the livestock trade in Europe. In contrast, a significant component of human cryptosporidiosis is traveller's diarrhoea – and even where contracted domestically, the source of infection is frequently distant[32,33,34]. We propose that the difference in migration patterns between the primary hosts can explain why we find no evidence of isolation-by-distance for *C. p. anthroponosum* in Europe, whilst there is strong geographic structuring in *C. p. parvum.* Differences in the rate of gene flow can also explain the notable distinction in the nucleotide diversity between these subspecies, which is almost an order of magnitude higher in *C. p. anthroponosum* than in *C. p. parvum*. Interestingly, parasite species from the *Plasmodium* genus show the opposite pattern in that the human-infective parasite species (*P. falciparum* and *P. malariae*) have a significantly lower nucleotide diversity compared to related zoonotic malarias (*P. reichenowi* and *P. malariae*-like)[35,36]. In this example, the lack of diversity in human-infective species has been interpreted as evidence for their recent population expansions. In *C. p. anthroponosum*, however, our population genetic analysis suggests that nucleotide diversity in the European population has been restored by introduction of novel genetic variation through immigration from diverged source populations outside Europe, as well as by genetic introgression.

Besides gene flow, our analysis identifies a strong signal of hybridisation between diverged strains or species, and we suggest that such genetic exchange between diverged taxa (i.e. genetic introgression) may also have contributed to the rapid restoration of diversity of *C. p. anthroponosum.* We detect 104 unique recombination events and estimate that the genetic exchanges have taken place relatively recently, i.e. within the last millennium or ~100,000 generations. This implies that hybridisation plays an important role in the biology of *Cryptosporidium*, and that this complex of *Cryptosporidium* species is coevolving in the presence of recent or continued genetic exchange. This interpretation is consistent with the growing body of evidence suggesting that hybridisation of diverged strains plays an important role in pathogen evolution[6,37]. Hybridisation can lead to the sharing of conserved

haploblocks across distinct phylogenetic lineages or (sub)species. Such mosaic-like genomes have been observed also in other human pathogens like *Toxoplasma gondii*[7], as well some plant pathogens such as the oomycete, *Albugo candida*[38]. Hybridisation can only occur, however, when different strains are in physical contact with one another. Unlike *A. candida*, which appears to suppress the host's immune response and facilitate coinfections[38], challenge experiments with human-infective isolates have shown that different *Cryptosporidium* species compete with each other within the host. For example, the *C. parvum parvum* strain GCH1 (subtype IIa) was shown to rapidly outcompete *C. hominis* strain TU502 (subtype Ia) during mixed infections in piglets[39]. Nevertheless, mixed species infections or intra-species diversity in *Cryptosporidium* have been identified in a large number (n = 55) of epidemiological surveys of cryptosporidiosis conducted in the period between 2005 – 2015[40]. As with *A. candida*, during the potentially brief periods of coinfections, hybridisation between distinct *Cryptosporidium* lineages may take place within a single host. In turn, this could facilitate the genetic exchange between the diverged lineages and contribute to the (virulence) evolution of *Cryptosporidium*. Introgression from an unidentified source into chromosome 8 of isolate UKP16 illustrates the diversity of the genepool that is able to exchange genetic variation, and it highlights the need for whole genome sequence studies for our understanding of *Cryptosporidium* biology. Interestingly, the distribution of recombination events varies markedly across chromosomes, a pattern observed also in other pathogens such as *T. gondii*[7]. Most remarkably, however, we found that in *Cryptosporidium* genes with a signature of positive selection were significantly more likely to be located in recombination blocks than neutrally evolving genes and genes under purifying selection. Our analyses thus suggest that such exchange is unlikely to be a neutral process, and that the recent emergence of the specialised anthroponotic subspecies such as *C. p. anthroponosum* might be fuelled by relatively recent, and possibly ongoing, "adaptive introgression"[37]. We estimate that these founding introgression events in the divergence of zoonotic *C. p. parvum* from the anthroponotic *C. p. anthroponosum* began 55-164 years ago, whereas those between *C. hominis* and *C. parvum* occurred between 410-1096 years ago timing which is consistent with reduced livestock contact and increased human population densities – conditions providing a continued selection pressure for the emergence of new human adapted pathogens from zoonotic origins.

Methods

Systematic Review

A human cryptosporidiosis prevalence database was constructed using data from peer-reviewed publications retrieved using the search term "Cryptosporidium" from PubMed (https://www.ncbi.nlm.nih.gov/pubmed) published between 2000-2017. After filtering (see SI Methods), the final dataset consisted of 743 publications of human *Cryptosporidium* infections in 126 countries.

Empirical Data

Whole genome sequence (WGS) data for *C. hominis* UKH1 and *C. meleagridis* UKMEL 1 were retrieved from the *Cryptosporidium* genetics database resource CryptoDB (www.cryptoDB.org)[41]. The remaining 19 *Cryptosporidium spp*. WGS datasets were obtained from clinical isolates[8] (see Table S1).

Concatenated Phylogenetic Analysis

61 neutrally-evolving loci (Ka/Ks = 0.2-0.6; 93.0-98.0% nucleotide IDs) between *C. parvum* UKP6 and *C. hominis* UKH4 were concatenated. A concatenated approach targeting neutral loci was used in lieu of the well-known gp60 subtyping locus, as this highly recombinant locus frequently produces phylogenies that do not correlate with genome-wide divergence (Fig. S7)[42]. Orthologous protein coding sequences from the human-infective WGS UKP6 and UKH4 were extracted (Table S10), and aligned using ClustalW. The Maximum Likelihood phylogeny was constructed with the Dayhoff substitution model, the Nearest-Neighbour-Interchange method and 2,000 bootstraps[43]. Divergence statistics between lineages were calculated using MEGA7[43].

Whole Genome Comparisons

Parallel whole genome comparative analyses were performed between a zoonotic *C. p. parvum* IIaA15G2R1-subtype WGS (UKP6), anthroponotic *C. p. anthroponosum* IIcA5G3a-subtype (UKP15), and anthroponotic *C. hominis* IaA14R3-subtype (UKH4). CDS nucleotide divergence was evaluated by cross-blasting CDS datasets locally (BLOSUM62 substitution matrix; BioEdit)[44]. Amino acid identities and indels resulting in frameshift were identified using EMBOSS Stretcher[45]. Selection was identified by calculating Ka/Ks in CodeML of PAML[46], and NaturalSelection.jl (https://github.com/BioJulia/NaturalSelection.jl). Sliding window Ka/Ks analyses, indel characterisations, and F_{ST} calculations were performed in DnaSP 5.10.1[47]. Putative protein function was evaluated using the UniProt BLASTp function (cut-off E-value <10e-5)[48], and putative protein localization was estimated using WoLF PSORT[49].

Phylogenomic analysis

Sequence reads of 21 *Cryptosporidium* isolates (Table S1) were aligned to the *C. parvum* Iowa[9] reference genome and SNPs identified (see SI Methods). Pseudoreferences were generated with filtered biallelic SNPs inserted using GATK FastaAlternateReferenceMaker[50]. Principle component analysis of *C. p. parvum* and *C. p. anthroponosum* isolates was performed with SNPrelate[51]. Population genetic statistics the fixation index (F_{ST}), absolute divergence (d_{xy}) and nucleotide diversity (π) were estimated in 50 Kb sliding windows (10 Kb step size) across the genome. Maximum likelihood phylogenies were estimated for 50 SNP windows across the genome using RAxML[52]. Topology weighting[23] was used to investigate the distribution of phylogenetic relationships across the genome with each isolate assigned to one of four groups (*C. p. parvum*, *C. p. anthroponosum*, UKP16 and outgroup samples (*C. hominis* and *C. cuniculus*). Ultrametric phylogenetic trees were made using the *chronopl* function in APE[53], and a consensus phylogeny was generated.

Recombination Analysis

Recombination signals due to introgression were detected using RDP4[54]. Automated detection algorithms RDP, GENECONV, Bootscan, Maxchi, and Chimaera were run with default values. Alternative call (AC) values of all bases in the four isolates that were studied in the genetic introgression analysis (UKH1, UKP6, UKP15 and UKP16) to validate that they comprised single subtype infections (Fig. S8).

Dating introgression events

Hybridization dating was estimated for introgressed regions in HybridCheck[55]. The HKY85 substitution model with a SNP mutation rate of $\mu=10^{-8}$ per generation was assumed, based on the observed nucleotide divergence between two outbreak WGS sampled seven days apart (Table S8). To convert generations into time, we assumed a factor of 12 autoinfective offspring per parental oocyst *in vivo* (Fig. S9). Furthermore, past infectivity studies revealed a population expansion of 3-5 new generations, and an estimated life cycle duration of 48-96h per infection (Table S9)[60,61]. This estimate is longer than previous estimates (12-14h)[56], but consistent with estimates of 72h from a cell culture experiment[57]. The reported estimates of time may be underestimated if oocysts remain dormant in the environment between infections of different host individuals.

Population Genetic Analysis

A total of 467 gp60 sequences collected in 43 countries were used to analyse the population structure of *C. p. parvum* UKP6 (N=361) and *C. p. anthroponosum* UKP15 (N=106) (see SI Methods). Population genetic structure was visualised using Fluxus network using median joining setting[58]. Isolation-by-distance analysis was performed using a regression analysis of the genetic distance (Kxy) between isolates and geographic distance between the sampling locations. Differences between chromosomes, chromosomal regions, recombinant regions and genes in the number of SNPs, indels, and recombination events were tested with Chi-square and binomial tests. Differences in nucleotide substitution patterns, indels and recombination events between taxa were analysed using Mann-Whitney test and ANOVAs. All tests were conducted in R (R Core Team)[59] and Minitab 12.1.

Data availability

All WGS data used in this paper is available publically and for free via the NCBI server (https://www.ncbi.nlm.nih.gov/) or CryptoDB (http://cryptodb.org/cryptodb/). The accession codes for the data are provided in Table S1.

Author's contributions

KT, RC, PH, JN and CvO conceived the study. JN and CvO designed the analyses. JN, JP, GR, MS, PH, KT and RC were involved in the acquisition of data. JN conducted the meta-analysis. JN and CvO conducted the evolutionary genetic analyses with input of TM for the phylogenetic and BW for the recombinant analyses. JN and CvO drafted the submitted manuscript. All authors contributed to revising the draft, had full access to all the data and read and approved the final manuscript.

Acknowledgements

This work was supported with funds awarded to KT and RMC from the FP7 KBBE EU project AQUAVALENS, grant agreement 311846 from the European Union awarded to PH, and a Biotechnology and Biological Sciences Research Council (BBSRC) (BB/N02317X/1) awarded to CvO, as well as support by the Earth & Life Systems Alliance (ELSA). P.R.H. is supported by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Gastrointestinal Infections at the University of Liverpool, in partnership with Public Health England (PHE), and in collaboration with University of East Anglia, University of Oxford, and the Institute of Food Research. Professor Hunter is based at the University of East Anglia. The views expressed are those of the authors and not necessarily those of the National Health Service, the National Institute for Health Research, the Department of Health, or Public Health England. We thank Gregorio Pérez-Cordóna for VNTR validation of isolates, and we thank the three reviewers for their helpful comments.

Competing Interests

The authors declare that there is no conflict of interest regarding the publication of this article.

References

[1]Liu, L. *et al*. Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. *Lancet* **379**, 2151-2161 (2012).

[2]Kotloff, K. L. *et al*. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *Lancet* **382**, 209-222 (2013).

[3]Widmer, G., & Sullivan, S. Genomics and population biology of *Cryptosporidium* species. *Parasite Immunol.* **34**, 61-71 (2012).

[4]Mazurie, A. *et al*. Comparative genomics of *Cryptosporidium*. *Int. J. Genomics* **2013**, 832756 (2013).

[5]Bushell, E. *et al*. Functional profiling of a *Plasmodium* genome reveals an abundance of essential genes. *Cell* **170**, 260-72 (2017).

[6]McMullan, M. *et al*. Evidence for suppression of immunity as a driver for genomic introgressions and host range expansion in races of *Albugo candida*, a generalist parasite. *eLife* **4** (2015).

[7]Lorenzi, H. *et al*. Local admixture of amplified and diversified secreted pathogenesis determinants shapes mosaic *Toxoplasma gondii* genomes. *Nature Commun.* **7**, 10147 (2016).

[8]Hadfield, S. J. *et al*. Generation of whole genome sequences of new *Cryptosporidium hominis* and *Cryptosporidium parvum* isolates directly from stool samples. *BMC Genomics* **16**, 1-12 (2015).

[9]Abrahamsen, M. S. *et al*. Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science* **304**, 441-445 (2004).

[10]Xu, P. *et al*. The genome of *Cryptosporidium hominis*. *Nature* **431**, 1107-1112 (2004).

[11]Bouzid, M., Hunter, P. R., Chalmers, R. M., & Tyler, K. M. *Cryptosporidium* pathogenicity and virulence. *Clin. Microbiol. Rev.* **26**, 115-134 (2013).

[12]Widmer, G. *et al*. Comparative genome analysis of two *Cryptosporidium parvum* isolates with different host range. *Infect. Genet. Evol.* **12**, 1213-1221 (2012).

[13]Guo, Y. *et al*. Comparative genomic analysis reveals occurrence of genetic recombination in virulent *Cryptosporidium hominis* subtypes and telomeric gene duplications in *Cryptosporidium parvum*. *BMC Genomics* **16**, 1-18 (2015).

[14]Li, N. *et al*. Genetic recombination and *Cryptosporidium hominis* virulent subtype IbA10G2. *Emerg. Infect. Dis.* **19**, 1573-82 (2013).

[15]Xiao, L. & Ryan U. M. Cryptosporidiosis: an update in molecular epidemiology. *Curr. Opin. Infect. Dis.* **17**, 483-90 (2004).

[16]Puleston, R. L. *et al*. The first recorded outbreak of cryptosporidiosis due to *Cryptosporidium cuniculus* (formerly rabbit genotype), following a water quality incident. *J. Water Health* **12**, 41-50 (2014).

[17]Koehler, A. V., Whipp, M. J., Haydon, S. R. & Gasser, R. B. *Cryptosporidium cuniculus* - new records in human and kangaroo in Australia. *Parasit. Vectors* **7**, 492 (2014).

[18]Wang, Y. *et al*. Population genetics of *Cryptosporidium meleagridis* in humans and birds: evidence for cross-species transmission. *Int. J. Parasitol*. **44**, 515-21 (2014).

[19]Koehler, A. V. *et al*. *Cryptosporidium viatorum* from the native Australian swamp rat Rattus lutreolus - An emerging zoonotic pathogen? *Int. J. Parasitol. Parasites Wildl*. **7**, 18-26 (2018).

[20]Li, N. *et al*. Subtyping *Cryptosporidium ubiquitum*, a zoonotic pathogen emerging in humans. *Emerg. Infect. Dis.* **20**, 217-24 (2014).

[21]Joachim, A. Human cryptosporidiosis: an update with special emphasis on the situation in Europe. *J. Vet. Med. B Infect. Dis. Vet. Public Health* **51**, 251-9. (2004).

[22]Chappell, C. L. et al. *Cryptosporidium muris*: infectivity and illness in healthy adult volunteers. *Am. J. Trop. Med. Hyg*. **92**, 50-5 (2015).

[23]Martin, S. H. & Van Belleghem, S. M. Exploring evolutionary relationships across the genome using topology weighting. *Genetics* **206**, 429-438 (2017).

[24]Okhuysen, P. C. *et al*. Infectivity of a *Cryptosporidium parvum* isolate of cervine origin for healthy adults and interferon-gamma knockout mice. *J. Infect. Dis*. **185**, 1320-5 (2002).

[25]Chappell, C. L. *et al*. *Cryptosporidium meleagridis*: infectivity in healthy adult volunteers. *Am. J. Trop. Med. Hyg*. **85**, 238-42 (2011).

[26]Santín, M., Trout, J. M., & Fayer, R. A longitudinal study of cryptosporidiosis in dairy cattle from birth to 2 years of age. *Vet. Parasitol.* **155**, 15-23 (2008).

[27]Current, W. L. Cryptosporidiosis. *J. Am. Vet. Med. Assoc*. **187**, 1334-8 (1985).

[28]Animal Transport Guides, Transport of calves. (2017). at: <http://animaltransportguides.eu/>.

[29]Defra,. PB 12544a: Welfare of Animals During Transport. (2011).

[30]LAres, E., & Ward, M. Live animal exports. *Commons Library Briefing.* **8031** (2017).

[31]ONS. Travel Trends: 2016. (2017). <https://www.ons.gov.uk/peoplepopulationandcommunity/leisureandtourism/articles/traveltrends/2016>.

[32]Jelinek, T. *et al*. Prevalence of infection with *Cryptosporidium parvum* and *Cyclospora cayetanensis* among international travellers. *Gut* **41**, 801-804 (1997).

[33]Nair, P. *et al*. Epidemiology of cryptosporidiosis in North American travelers to Mexico. *Am. J. Trop. Med. Hyg.* **79**, 210-4 (2008).

[34]Chalmers, R. M. *et al*. Geographic linkage and variation in *Cryptosporidium hominis*. *Emerg. Infect. Dis.* **14**, 496-8 (2008).

[35]Sundararaman, S. A. *et al*. Genomes of cryptic chimpanzee *Plasmodium* species reveal key evolutionary events leading to human malaria. *Nat. Commun*. **22**, 11078 (2016).

[36]Rutledge, G. G. *et al*. *Plasmodium malariae* and *P. ovale* genomes provide insights into malaria parasite evolution. *Nature* **542**, 101-104 (2017).

[37]King, K. C., Stelkens, R. B., Webster, J. P., Smith, D. F. & Brockhurst, M. A. Hybridization in parasites: consequences for adaptive evolution, pathogenesis, and public health in a changing world. *PLoS Pathog.* **11** (2015).

[38]Jouet, A. *et al*. *Albugo candida* race diversity, ploidy and host□associated microbes revealed using DNA sequence capture on diseased plants in the field. *New Phytol.* doi: 10.1111/nph.15417 (2018).

[39]Akiyoshi, D. E., Mor, S. & Tzipori, S. Rapid displacement of *Cryptosporidium parvum* type 1 by type 2 in mixed infections in piglets. *Infect. Immun.* **71**, 5765-71 (2003).

[40]Grinberg, A. & Widmer, G. *Cryptosporidium* within-host genetic diversity: systematic bibliographical search and narrative overview. *Int. J. Parasitol*. **46**, 465-71 (2016).

[41]Puiu, D. *et al*. CryptoDB: the *Cryptosporidium* genome resource. *Nucleic Acids Res.* **32**, D329-31 (2004).

[42]Feng, Y., Ryan, U. M. & Xiao, L. Genetic diversity and population structure of *Cryptosporidium*. *Trends Parasitol.* **34**, 997-1011 (2018).

[43]Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol*. **33**, 1870-4 (2016).

[44]Hall, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**, 95-98 (1999).

[45]Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet*. **16**, 276-7 (2000).

[46]Suyama, M., Torrents, D. & Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res*. **34**, W609-12 (2006).

[47]Librado, P. & Rozas, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-2 (2009).

[48]Apweiler, R. *et al*. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res*. **32**, D115-9 (2004).

[49]Horton, P. *et al*. WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* **35**, W585-7 (2007).

[50]DePristo, M. A. *et al*. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genet.* **43**, 491-8 (2011).

548 [51]Zheng, X. *et al*. A High-performance Computing Toolset for Relatedness and Principal
549 Component Analysis of SNP Data. *Bioinformatics* **28**, 3326-3328 (2012).

550 [52]Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
551 large phylogenies. *Bioinformatics* **30**, 1312-3 (2014).

552 [53]Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R
553 language. *Bioinformatics* **20**, 289-90 (2004).

554 [54]Martin, D. P., Murrell, B., Golden, M., Khoosal, A. & Muhire, B. RDP4: Detection and
555 analysis of recombination patterns in virus genomes. *Virus Evol.* **1**, vev003 (2015).

556 [55]Ward, B. J. & van Oosterhout, C. HYBRIDCHECK: software for the rapid detection,
557 visualization and dating of recombinant regions in genome sequence data. *Mol. Ecol. Resour*.
558 **16**, 534-9 (2016).

559 [56]Fleming, R. *Cryptosporidium*: Could It Be in Your Water? Ontario Ministry of Agriculture,
560 Food, and Rural Affairs (2015).

561 [57]Current, W. L. & Haynes, T. B. Complete development of *Cryptosporidium* in cell culture.
562 *Science* **224**, 604-5 (1984).

563 [58]Bandelt, H. J., Forster, P. & Rohl, A. Median-joining networks for inferring intraspecific
564 phylogenies. *Mol. Biol. Evol*. **16**, 37–48 (1999).

565 [59]R Core Team. R: A language and environment for statistical computing. R Foundation for
566 Statistical Computing, Vienna, Austria (2013). URL http://www.R-project.org/.

567 [60]Kosek, M., Alcantara, C., Lima, A. A. & Guerrant, R. L. Cryptosporidiosis: an update.
568 *Lancet Infect. Dis*. **1**, 262-9 (2001).
569
570 [61]O'Hara, S. P. & Chen, X. M. The cell biology of *Cryptosporidium* infection. *Microbes*
571 *Infect*. **13**, 721-30 (2011).

572

Legends to Figures

Figure 1

a, Concatenated phylogeny of 16 human-infective *Cryptosporidium spp*. The maximum likelihood phylogeny is based on a 142,452 bp alignment of 61 loci (Table S10) and 2,000 bootstrap replications. Unique UK-identifiers show species group, specific gp60 subtype, and prevalent host type(s) (Table S1, Fig. S1). **b,c,** Relative global distribution of human cryptosporidiosis due to *C. parvum* (orange) versus *C. hominis* (blue) based on a systematic review of 743 peer-reviewed publications (Dropbox). Relative proportion of global *C. parvum* human cryptosporidiosis due to zoonotic *C. p. parvum* IIa (green) versus anthroponotic *C. p. anthroponosum* IIc-a (purple) based on a systematic review of 84 peer-reviewed publications. **d,** Nucleotide diversity (π) within European *C. p. parvum* (IIa) (green, n=96; Min=0.000000, 1st Qu.=0.001374, Median=0.002762, Mean=0.003244, 3rd Qu.=0.004169, Max=0.006970) and *C. p. anthroponosum* (IIc-a) (purple, n=22; Min=0.000000, 1st Qu.=0.002124, Median=0.043951, Mean=0.029704, 3rd Qu.=0.046250, Max=0.061045) populations. **e**, The genetic distance (Kxy) between *C. p. parvum* (n=345) isolates is strongly correlated with geographic distance (Regression $F_{1,26}$=40.63, p=0.000000944, R^2=61.0%), whilst there is no isolation-by-distance signal detected for *C. p. anthroponosum* (n=106) isolates ($F_{1,16}$=1.477, p=0.242). **f**, *C. p. parvum* (IIa) isolates show an isolation-by-distance signal, as is illustrated by the positive slope of the regression line between genetic differentiation (Fst) and geographic distance (Regression: R^2-adj.=58.3%, $F_{1,8}$=13.60, p=0.006). This signal suggests there is some gene flow within Europe. No isolation-by-distance was found for *C. p. anthroponosum* (IIc-a) in Europe. Combined with significantly higher nucleotide diversity, this suggests that *C. p. anthroponosum* infections arrive from outside Europe, rather than being transmitted within Europe. **g,h,** Fluxus network of global *C. p. parvum* (IIa) and *C. p. anthroponosum* (IIc-a) GenBank-submitted gp60 sequences show significant sub-structuring of global populations of *C. p. parvum* IIa isolates, and absence of structure between or within regional populations of *C. p. anthroponosum* IIc-a.

Figure 2

a,**b**, Selective pressures (Ka/Ks) and nucleotide distances (π) generated gene-by-gene between and within zoonotic and anthroponotic *Cryptosporidium* species groups. Zoonotic *C. p. parvum* UKP6 genomics coding sequences (CDSs) are here compared to zoonotic *C. p. parvum* UKP8 (green; Min=0.00000, 1st Qu.=0.00000, Median=0.00000, Mean=0.1613, 3rd Qu.=0.00000, Max=1.00000), anthroponotic *C. parvum parvum* UKP16 (yellow; Min=0.00000, 1st Qu.=0.00000, Median=0.00000, Mean=0.17991, 3rd Qu.=0.09046, Max=1.00000), anthroponotic *C. p. anthroponosum* UKP15 (red; Min=0.00000, 1st Qu.= 0.00000, Median=0.00000, Mean=0.2169, 3rd Qu.=0.2219, Max=1.00000), and anthroponotic *C. hominis* UKH4 (blue; Min=0.00000, 1st Qu.=0.05924, Median=0.11785, Mean=0.13858, 3rd Qu.=0.18854, Max=1.00000). Distribution of global Ka/(Ka+Ks) values for each comparison are shown, and differences were assessed statistically (One-way ANOVA, $F_{12,727} = 31.34$, P<3.567e-20, n=3465 CDSs). **c**, Sliding window analysis of triplet (brown) and non-triplet (green) insertion and deletion (indel) events between two samples, i.e. *C. parvum parvum* UKP6 and *C. parvum anthroponosum* UKP15. Composite results for 20 kb-wide sliding windows across chromosomes 1, 2, 4, 6, and 8 are shown. Peri-telomeric genes (T) and subtelomeric genes (S) have significantly more triplet and non-triplet indels than non-telomeric (NT) genes (Chi-sq. test, X^2=38.535, df=2, p=4.29x10^{-9}; X^2=226.078, df=2, p=8.09e^{-50}, respectively). **d**, Comparative selective pressure analysis between *C. p. parvum* UKP6 and *C. p. anthroponosum* UKP15 coding sequences with contrasting protein localizations. The range of Ka/(Ka+Ks) between all (n=3465; Min=0.00000, 1st Qu.=0.00000, Median=0.1416, Mean=0.3058, 3rd Qu.=0.3989, Max=1.00000) CDSs, CDSs annotated as having a cytoplasmic protein localization (n=1152; Min=0.00000, 1st Qu.=0.00000, Median=0.1110, Mean=0.2980, 3rd Qu.=0.3705, Max=1.00000), and CDSs annotated as having an extracellular localization (n=333; Min=0.00000, 1st Qu.=0.00000, Median=0.1973, Mean=0.4180, 3rd Qu.= 1.00000, Max=1.00000) are represented by a violin plot. CDSs with extracellular localisation experience significantly more positive selection than cytoplasmic CDSs, as evidenced by their higher Ka/(Ka+Ks) value (two-sided Mann-Whitney test, W=842985, p=0.0182). In addition, 17 out of 333 (5.1%) extracellular CDSs have a Ka/Ks larger than unity, compared to just 21 out of 3236 (0.6%) cytoplasmic CDSs (Chi-sq. test: X^2=53.8, d.f.=1, p=1.675e-12).

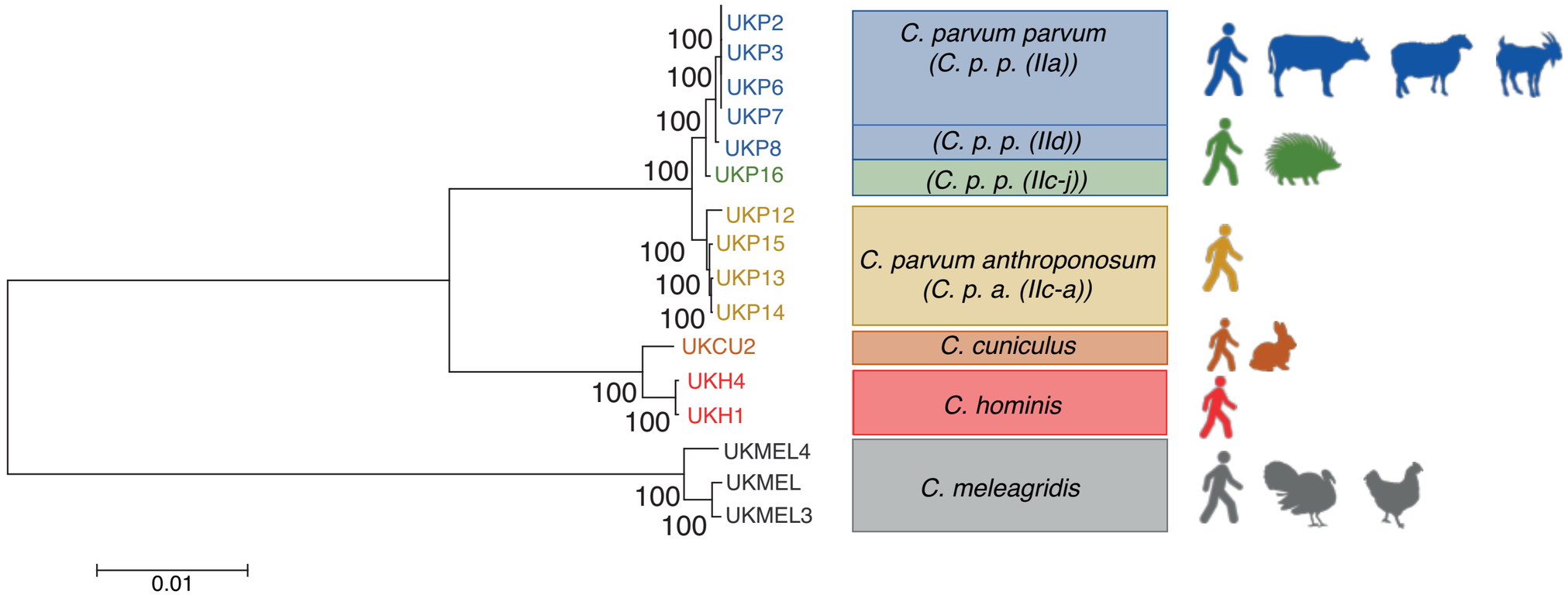
Figure 3

a, Principle component analysis of *C. p. parvum* and *C. p. anthroponosum* isolates based on 1,476 high quality SNPs retained after pruning based on linkage disequilibrium. **b**, A "cloudogram" of 1,324 trees showing phylogenomic relationships between WGS of anthroponotic *Cryptosporidium* isolates. Maximum likelihood trees were estimated for non-overlapping 50 SNP genomic windows across the *C. parvum* Iowa II reference genome (grey). The consensus phylogeny is shown in black. Isolates belonging to *C. p. parvum* and *C. p. anthroponosum* sub-species fall into two monophyletic groups, C. hominis/C.cuniculus isolates are included as an outgroup (OG). **c**, Topology weighting was used to explore the genome-wide distribution of phylogenetic relationships between the two *C. parvum* subspecies, a putatively introgressed isolate (UKP16) and an outgroup (*C. hominis* isolates and a single *C. cuniculus* isolate) using the 50 SNP fixed window trees. All possible topologies of the ingroup taxa are shown in the top panel, the lower panel shows the genome-wide average weighting of each topology. **d**, The distribution of topology weightings across chromosome 8 (colours as per c) reveals a putatively introgressed region between 500Kb and 650Kb. **e**, Absolute divergence (d_{xy}) between *Cryptosporidium* sub-species and the putatively introgressed isolate UKP16 in 50 Kb sliding windows (10Kb step size) across chromosome 8 of the *C. parvum* Iowa II reference genome.

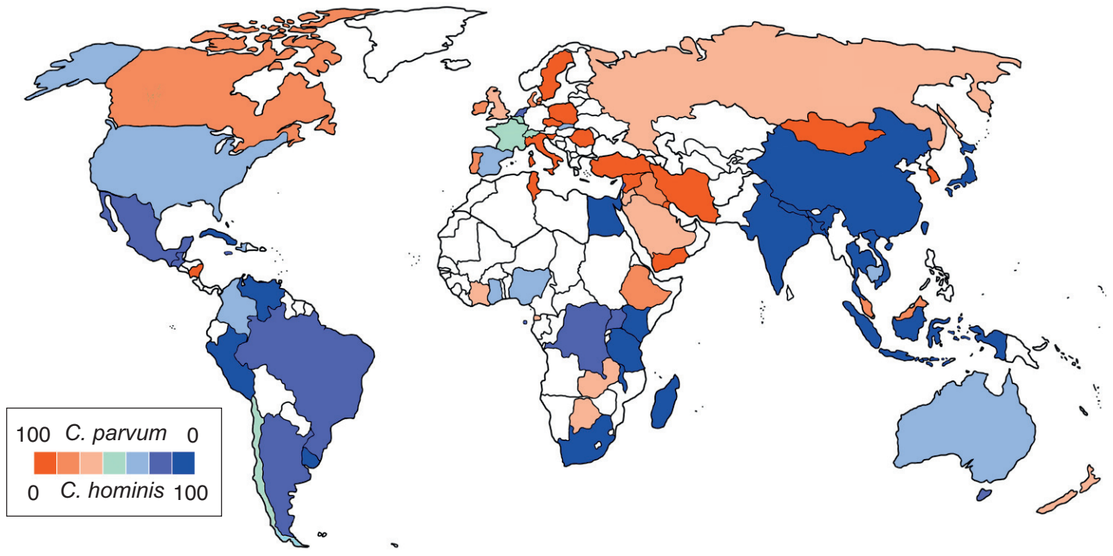
Figure 4

a, Genomic recombinant events in anthroponotic *Cryptosporidium spp*. WGS. Size and location of recombinant fragments detected by RDP4 are illustrated for recombination between *C. p. parvum* UKP6 and *C. p. parvum* UKP16 (yellow), *C. p. parvum* UKP6 and *C. p. anthroponosum* UKP15 (pink), *C. p. parvum* UKP16 and *C. p. anthroponosum* UKP15 (turquoise), *C. p. parvum* UKP6 and *C. hominis* UKH1 (green), *C. p. anthroponosum* UKP15 and *C. hominis* UKH1 (blue), and *C. p. parvum* UKP16 and *C. hominis* UKH1 (peach). Recombination events with unknown major or minor parentage are additionally represented (grey). Individual recombination events are detailed in Table S7. **b**, Estimated dates of introgression events between anthroponotic and zoonotic *Cryptosporidium spp*.. The range of estimated introgression times (thousands of generations ago) are given for introgression events between zoonotic *C. p. parvum* (UKP6) and anthroponotic *C. p. anthroponosum* (UKP15) – n=45, Min=7369, 1st Qu.=9218, Median=11486, 3rd Qu=13045, Max=17914 , and for introgression events between zoonotic *C. p. parvum* (UKP6) and anthroponotic *C. hominis* (UKH1) – n=33, Min=64655, 1st Qu.=77337, Median=95974, Mean=103281, 3rd Qu.117130, Max=188341. Minimum, mean, and maximum generation numbers were converted into units of time (years) for both 48- and 96-hour life cycle estimates.

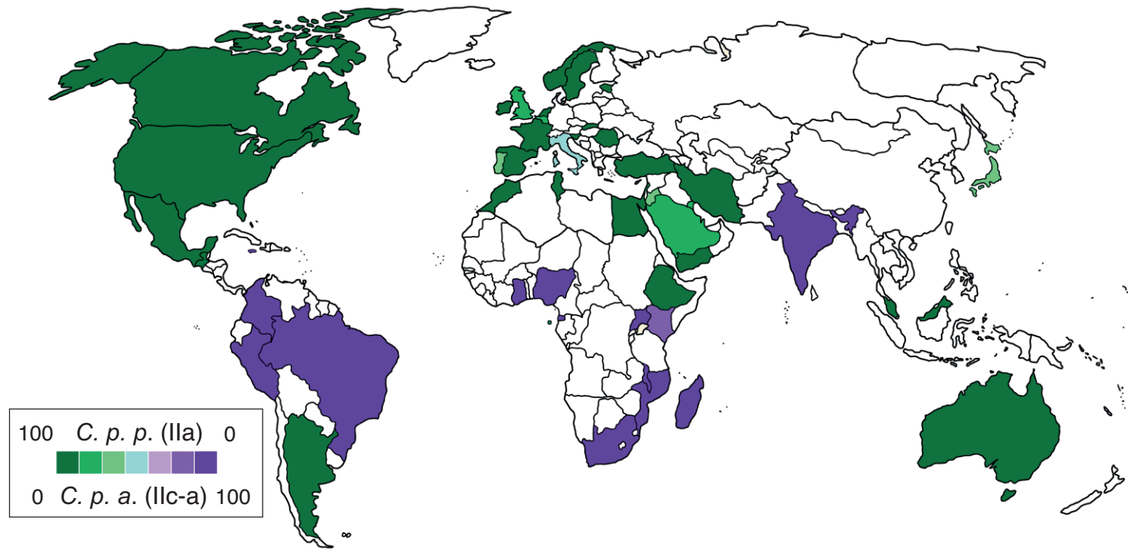
a


UKP2
UKP3
UKP6
UKP7
UKP8
UKP16
UKP12
UKP15
UKP13
UKP14
UKCU2
UKH4
UKH1
UKMEL4
UKMEL
UKMEL3
100
0.01
C. parvum parvum (C. p. p. (IIa))
(C. p. p. (IId))
(C. p. p. (IIc-j))
C. parvum anthroponosum (C. p. a. (IIc-a))
C. cuniculus
C. hominis
C. meleagridis

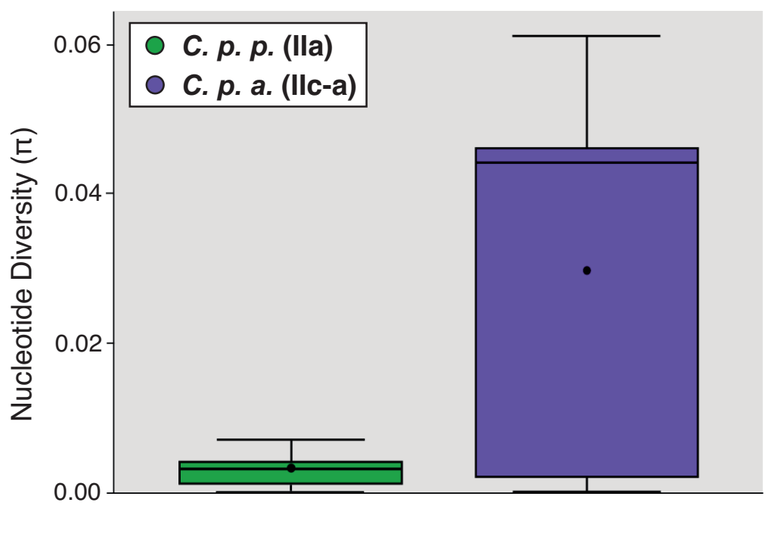
b


100 C. parvum 0
0 C. hominis 100

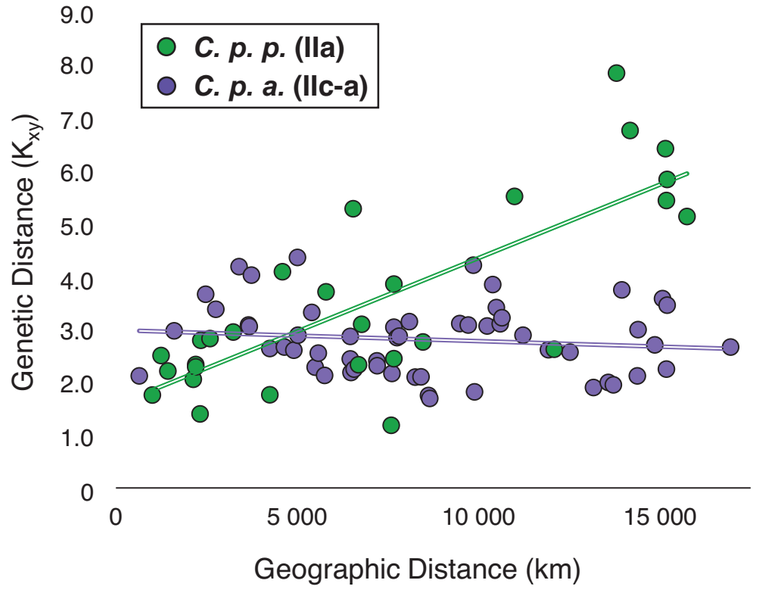
c


100 C. p. p. (IIa) 0
0 C. p. a. (IIc-a) 100

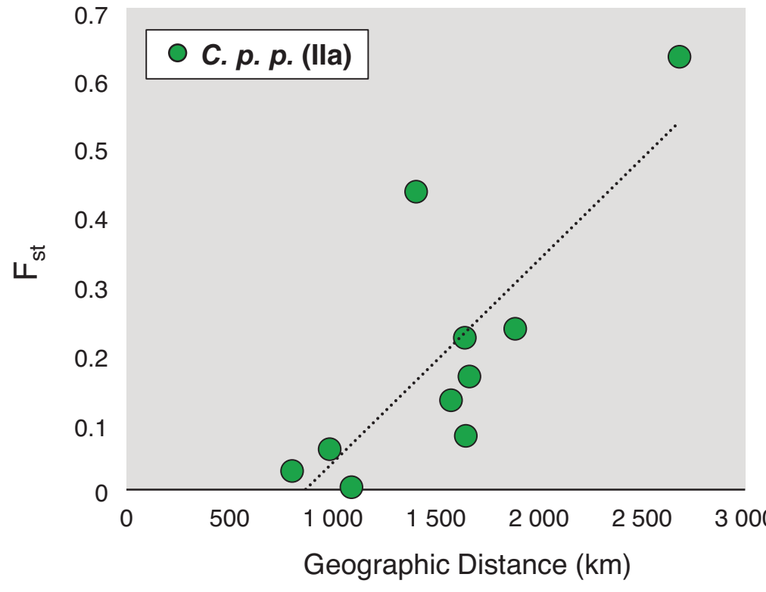
d


C. p. p. (IIa)
C. p. a. (IIc-a)
Nucleotide Diversity (π)
0.06
0.04
0.02
0.00

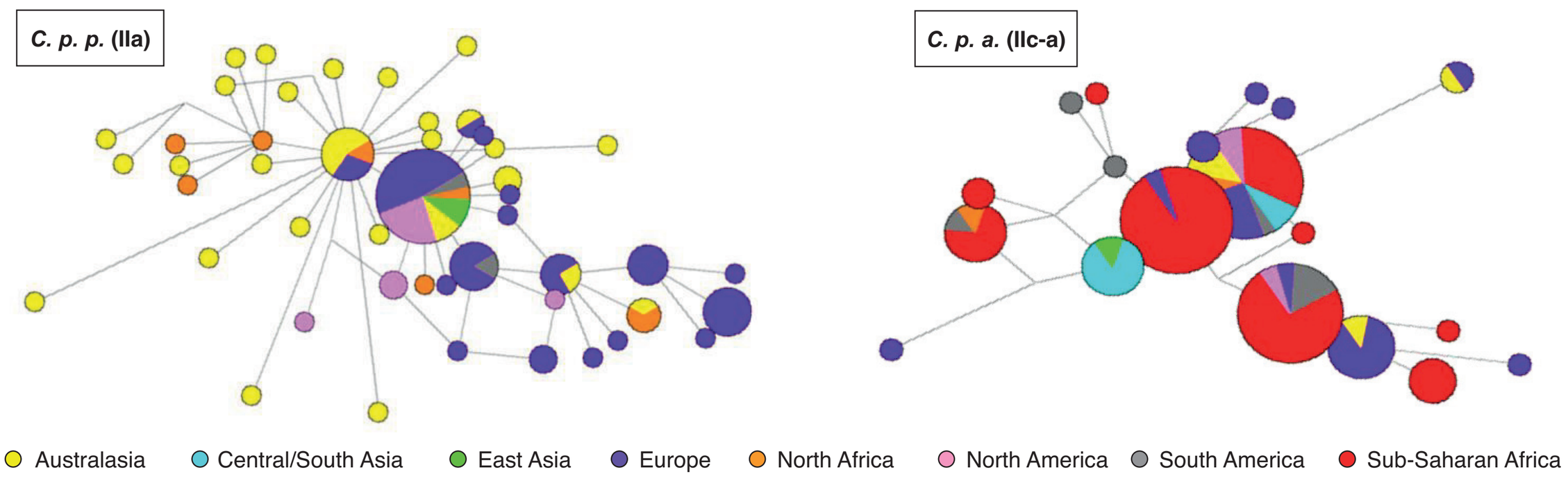
e


C. p. p. (IIa)
C. p. a. (IIc-a)
Genetic Distance (Kxy)
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0
5 000
10 000
15 000
Geographic Distance (km)

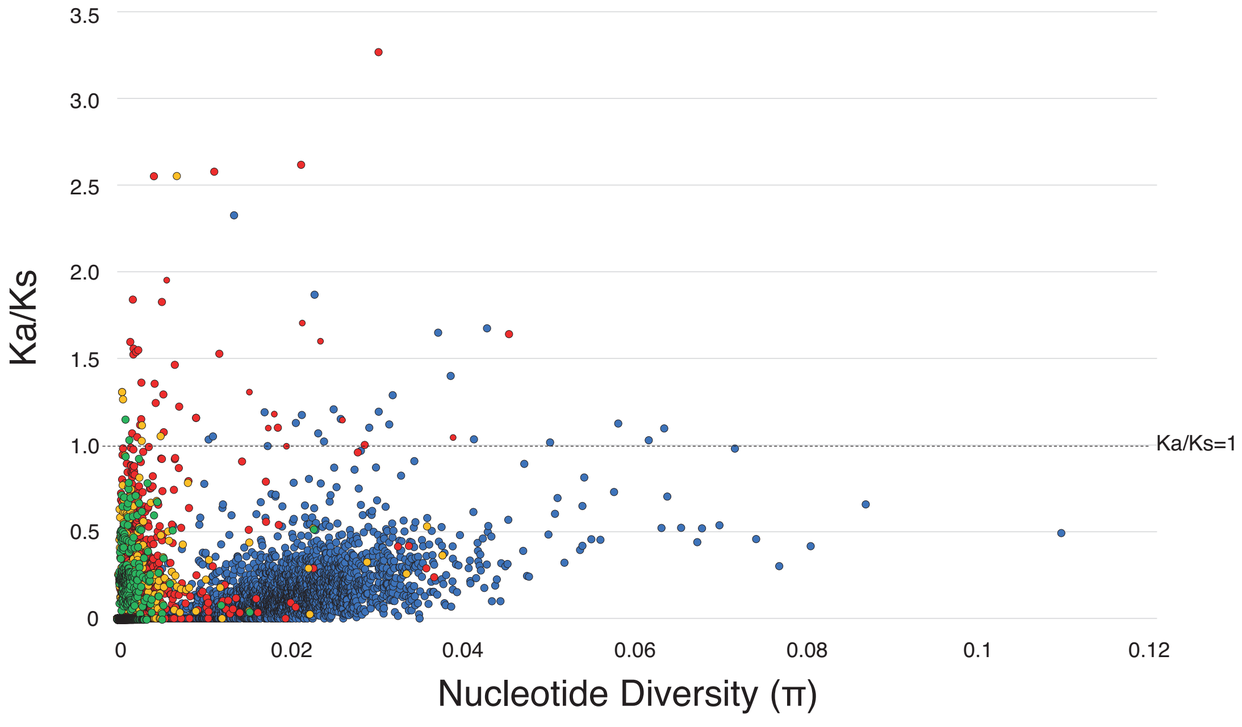
f


C. p. p. (IIa)
Fst
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
500
1 000
1 500
2 000
2 500
3 000
Geographic Distance (km)

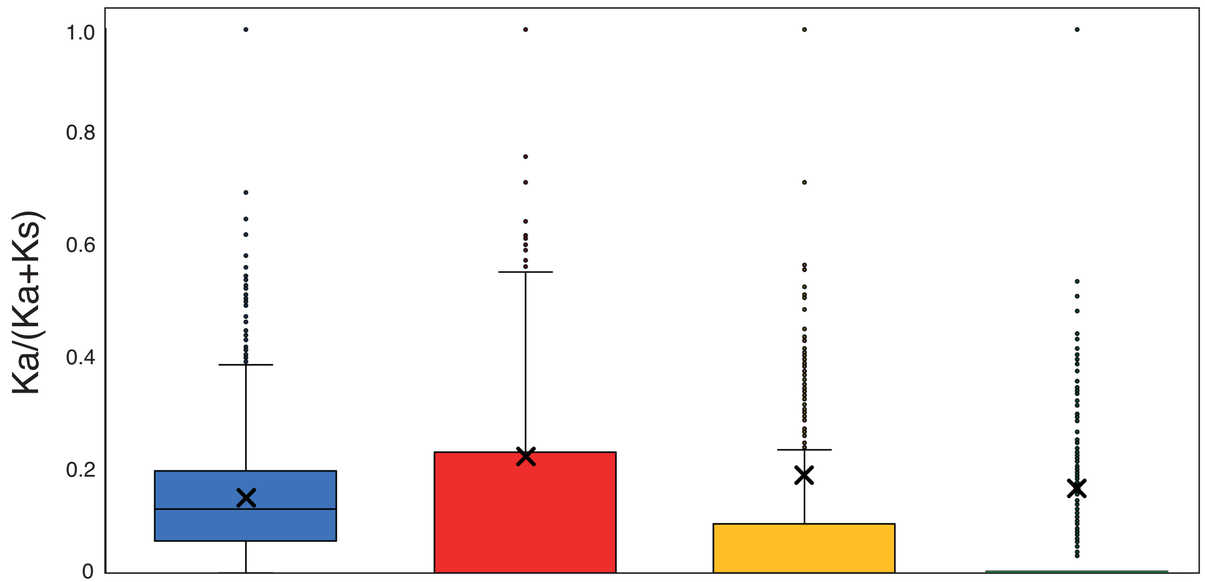
g


C. p. p. (IIa)
C. p. a. (IIc-a)
Australasia
Central/South Asia
East Asia
Europe
North Africa
North America
South America
Sub-Saharan Africa

a



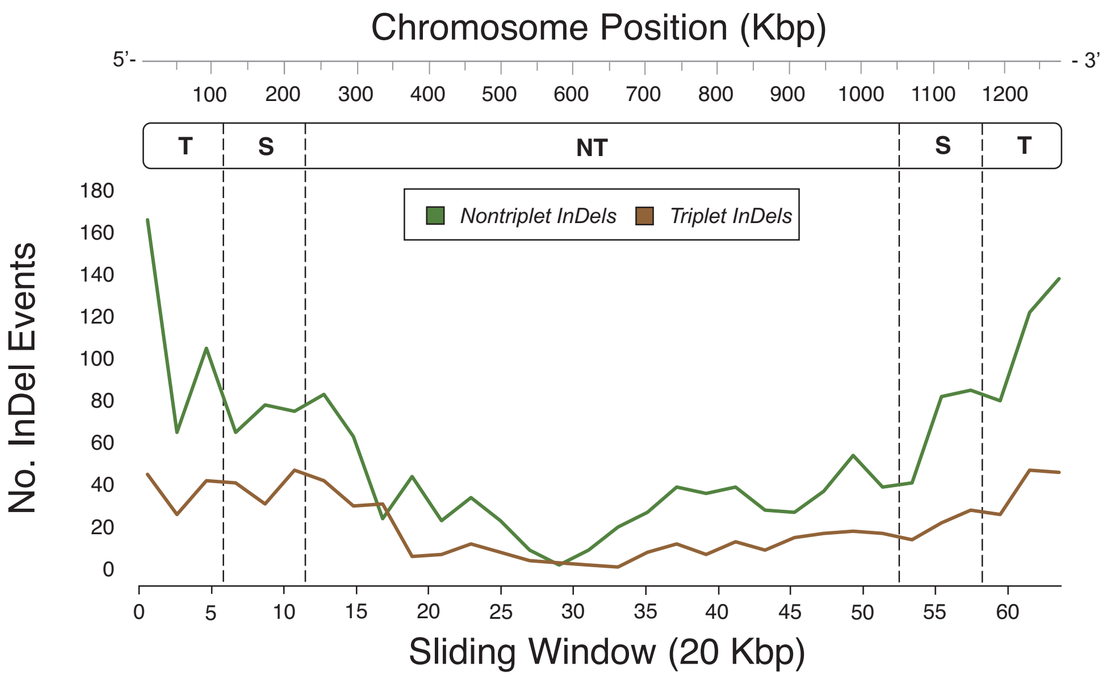
b



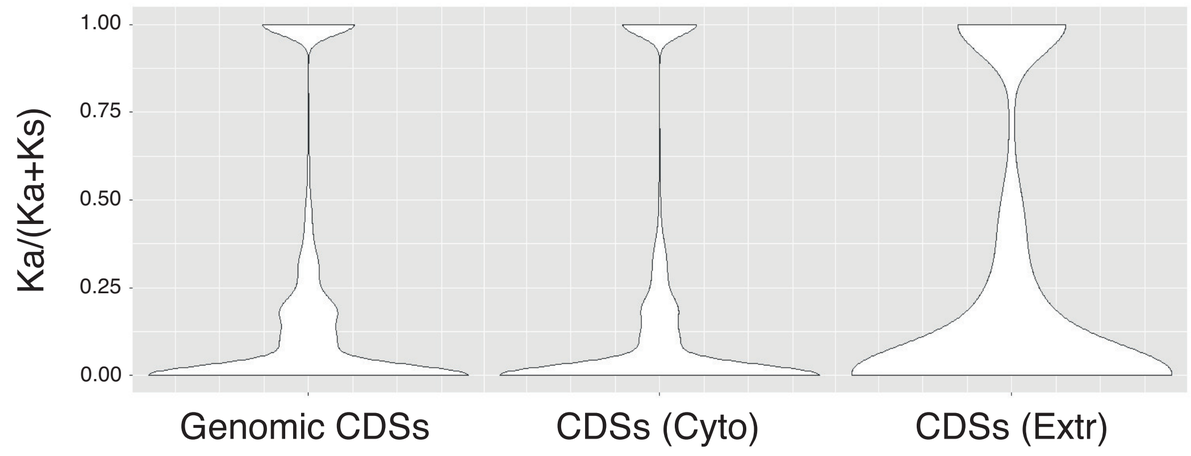
C. p. parvum UKP6 v. *C. hominis* UKH4
C. p. parvum UKP6 v. *C. p. parvum* UKP16
C. p. parvum UKP6 v. *C. p. anthroponosum* UKP15
C. p. parvum UKP6 v. *C. p. parvum* UKP8

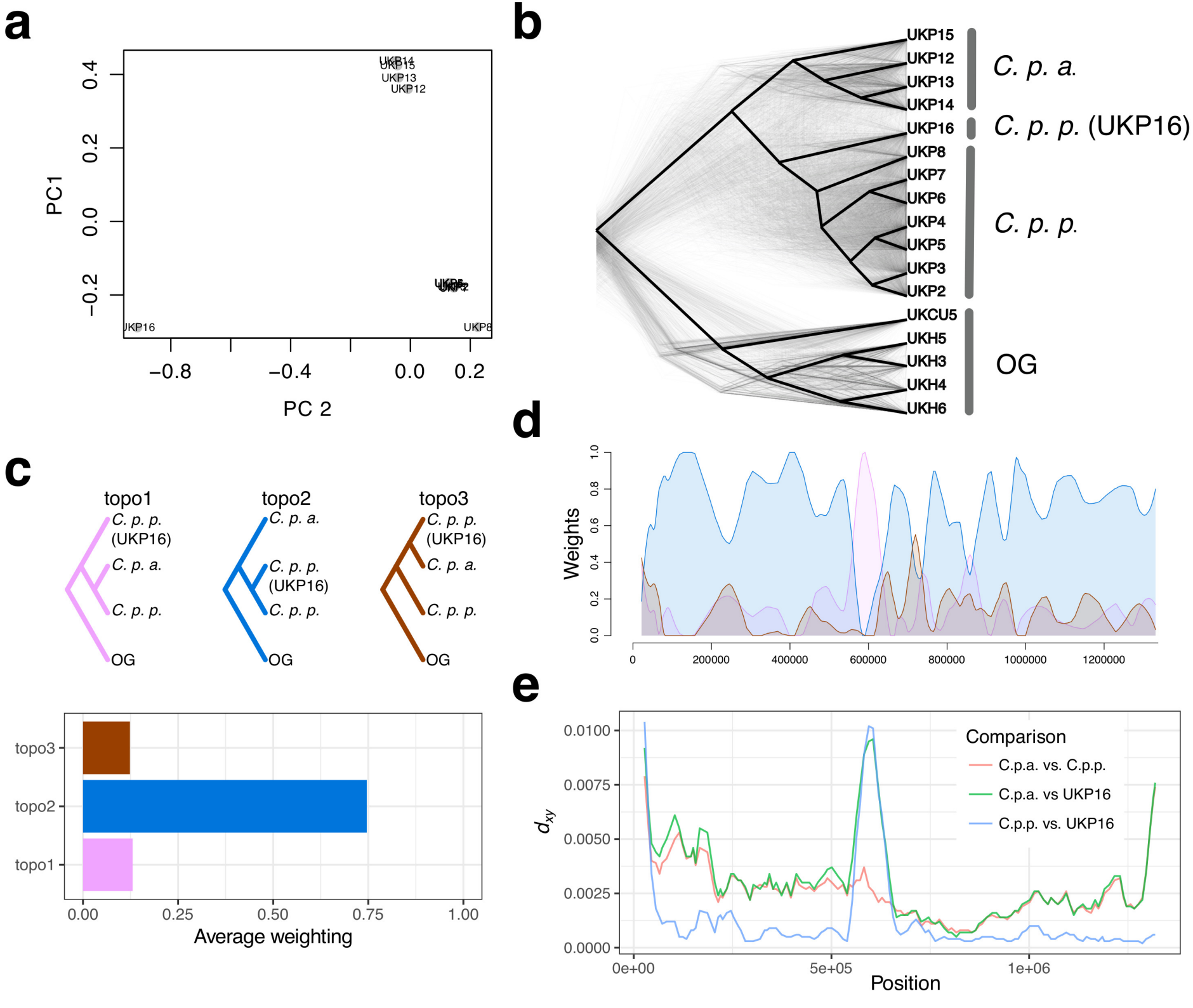
*ANOVA p<0.00001

c

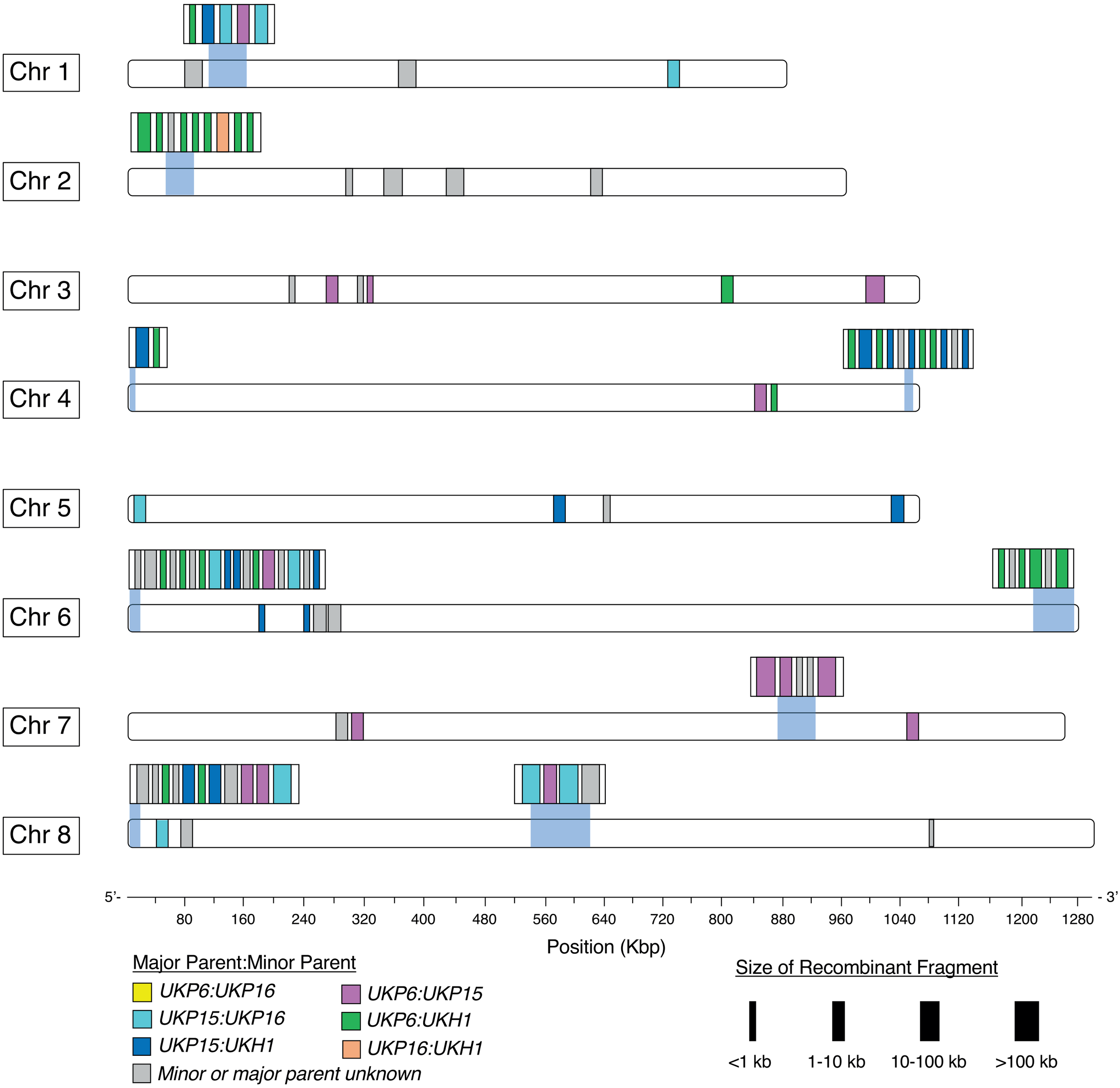


d

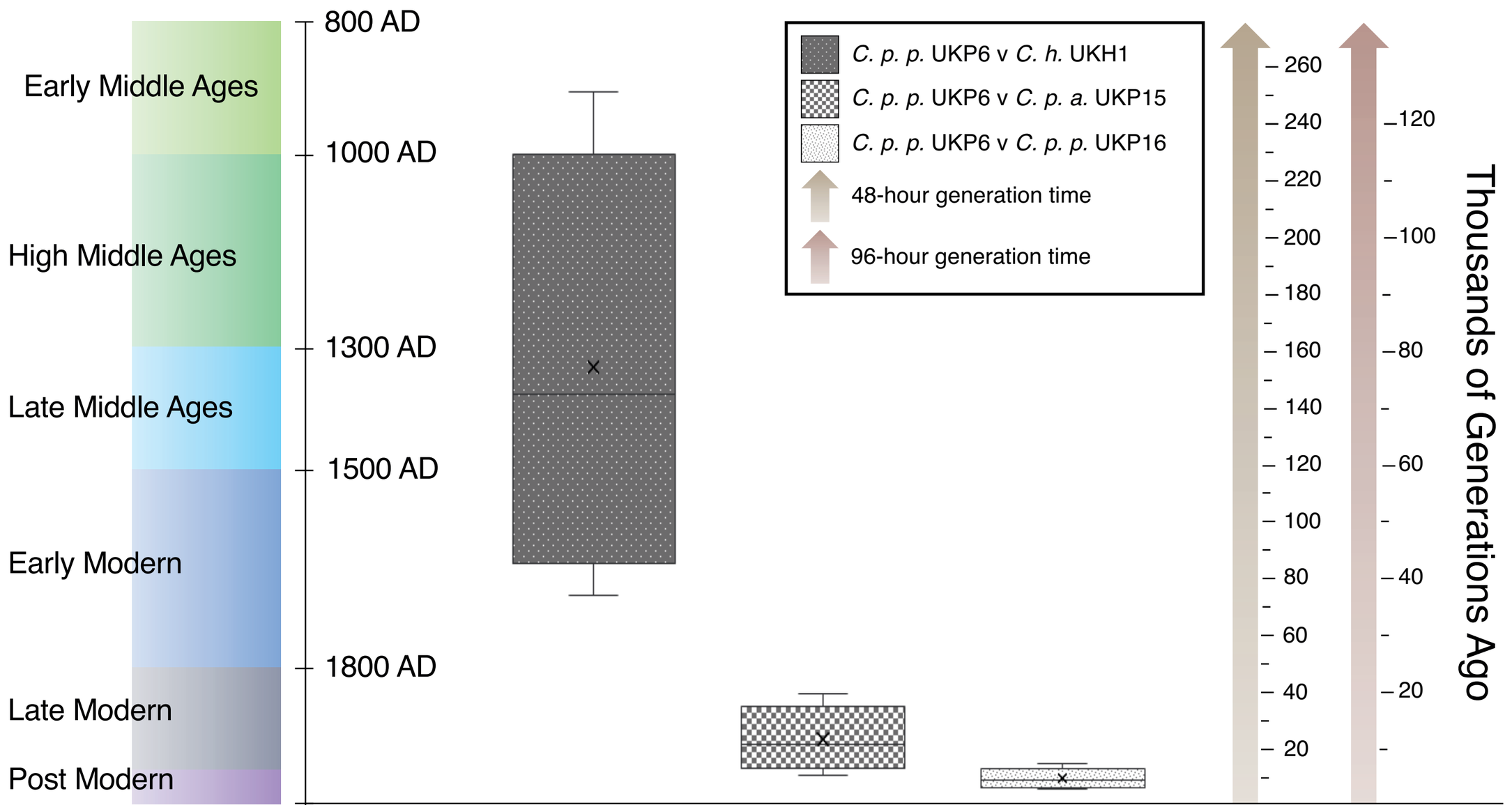



a
PC1
PC 2
0.4
0.2
0.0
−0.2
−0.8
−0.4
0.0
0.2
UKP14
UKP15
UKP13
UKP12
UKP16
UKP8
b
UKP15
UKP12
UKP13
UKP14
UKP16
UKP8
UKP7
UKP6
UKP4
UKP5
UKP3
UKP2
UKCU5
UKH5
UKH3
UKH4
UKH6
C. p. a.
C. p. p. (UKP16)
C. p. p.
OG
c
topo1
C. p. p. (UKP16)
C. p. a.
C. p. p.
OG
topo2
C. p. a.
C. p. p. (UKP16)
C. p. p.
OG
topo3
C. p. p. (UKP16)
C. p. a.
C. p. p.
OG
topo3
topo2
topo1
0.00
0.25
0.50
0.75
1.00
Average weighting
d
Weights
0.0
0.2
0.4
0.6
0.8
1.0
0
200000
400000
600000
800000
1000000
1200000
e
d_{xy}
0.0100
0.0075
0.0050
0.0025
0.0000
0e+00
5e+05
1e+06
Position
Comparison
C.p.a. vs. C.p.p.
C.p.a. vs UKP16
C.p.p. vs. UKP16

a


Chr 1
Chr 2
Chr 3
Chr 4
Chr 5
Chr 6
Chr 7
Chr 8
5'-
- 3'
80
160
240
320
400
480
560
640
720
800
880
960
1040
1120
1200
1280
Position (Kbp)
Major Parent:Minor Parent
UKP6:UKP16
UKP15:UKP16
UKP15:UKH1
UKP6:UKP15
UKP6:UKH1
UKP16:UKH1
Minor or major parent unknown
Size of Recombinant Fragment
<1 kb
1-10 kb
10-100 kb
>100 kb

b


Early Middle Ages
High Middle Ages
Late Middle Ages
Early Modern
Late Modern
Post Modern
800 AD
1000 AD
1300 AD
1500 AD
1800 AD
C. p. p. UKP6 v C. h. UKH1
C. p. p. UKP6 v C. p. a. UKP15
C. p. p. UKP6 v C. p. p. UKP16
48-hour generation time
96-hour generation time
260
240
220
200
180
160
140
120
100
80
60
40
20
120
100
80
60
40
20
Thousands of Generations Ago